Published online 17 November 2012 Nucleic Acids Research, 2013, Vol. 41, Database issue D177–D187
doi:10.1093/nar/gks1060

ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data

Jian-Hua Yang, Jun-Hao Li, Shan Jiang, Hui Zhou and Liang-Hu Qu*

RNA Information Center, Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory for Biocontrol, Sun Yat-sen University, Guangzhou 510275, P.R. China

Received August 15, 2012; Revised September 19, 2012; Accepted October 11, 2012

ABSTRACT

Long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) represent two classes of important non-coding RNAs in eukaryotes. Although these non-coding RNAs have been implicated in organismal development and in various human diseases, surprisingly little is known about their transcriptional regulation. Recent advances in chromatin immunoprecipitation with next-generation DNA sequencing (ChIP-Seq) have provided methods of detecting transcription factor binding sites (TFBSs) with unprecedented sensitivity. In this study, we describe ChIPBase (http://deepbase.sysu.edu.cn/chipbase/), a novel database that we have developed to facilitate the comprehensive annotation and discovery of transcription factor binding maps and transcriptional regulatory relationships of lncRNAs and miRNAs from ChIP-Seq data. The current release of ChIPBase includes high-throughput sequencing data that were generated by 543 ChIP-Seq experiments in diverse tissues and cell lines from six organisms. By analysing millions of TFBSs, we identified tens of thousands of TF-lncRNA and TF-miRNA regulatory relationships. Furthermore, two web-based servers were developed to annotate and discover transcriptional regulatory relationships of lncRNAs and miRNAs from ChIP-Seq data. In addition, we developed two genome browsers, deepView and genomeView, to provide integrated views of multidimensional data. Moreover, our web implementation supports diverse query types and the exploration of TFs, lncRNAs, miRNAs, gene ontologies and pathways.

INTRODUCTION

It has become increasingly clear that eukaryotic genomes encode thousands of long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) (1–4). Emerging evidence is revealing that lncRNAs and miRNAs serve as the nodes of signaling networks that regulate cancer, apoptosis, proliferation, differentiation and stem cell biology (1,2,5–8). However, the majority of studies that address these types of RNAs focus on defining the regulatory functions of lncRNAs and miRNAs, whereas few investigations are directed toward assessing how the lncRNA and miRNA genes themselves are transcriptionally regulated.

The major limitation in identifying the transcriptional regulatory relationships of lncRNAs and miRNAs has been the high false-positive rates of predictive algorithms for transcription factor binding sites (TFBSs) (9). Recently, chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-Seq) has provided a powerful way to identify TFBSs. The application of the ChIP-Seq technique has significantly reduced the rate of false-positive predictions of TFBSs (10–12). However, although ChIP-Seq technology can reliably identify TFBSs, few studies have used ChIP-Seq data to explore the transcriptional regulation of lncRNAs and miRNAs. For this reason, a high-quality integrated database that could facilitate the annotation and analysis of the transcriptional regulation of lncRNAs and miRNAs from ChIP-Seq data will be of great utility in the study of both the regulation of lncRNAs and miRNAs by TFs and the roles of this regulation in human diseases.

In this study, we developed the ChIPBase to facilitate the integrative and interactive display, as well as the comprehensive annotation and discovery, of TF-lncRNA and TF-miRNA interaction maps from ChIP-Seq data that were generated from diverse tissues and cell lines from six organisms: human, mouse, dog, chicken,

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

*To whom correspondence should be addressed. Tel: +86 20 84112399; Fax: +86 20 84036551; Email: lssqlh@mail.sysu.edu.cn

© The Author(s) 2012. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

Drosophila melanogaster and *Caenorhabditis elegans* (Figure 1). ChIPBase contains tens of thousands of TF-lncRNA and TF-miRNA regulatory relationships, as well as millions of TFBSs (Table 1). In addition, two novel web servers and two genome browsers were developed to comprehensively explore the relationships of TFs and ncRNAs from ChIP-Seq data.

MATERIALS AND METHODS

A total of 543 ChIP-Seq peak data sets for 252 different transcription factors were compiled from multiple related studies and downloaded from the NCBI GEO database (13), the ENCODE (14) and modENCODE (15,16) databases, or the Supplementary Data of the original research articles (Supplementary Table S1). We have also manually curated metadata (such as TF name, refSeq accession number, gene symbols and detailed descriptions and expression patterns of TFs) to ensure annotation consistency. These peak data sets were converted to latest genome version using liftOver tool from the UCSC genome browser website (17), and peaks whose genomic regions could not be transformed into latest version of the genome were discarded. In addition, some data sets in BedGraph format were downloaded to construct peak tracks and displayed in our deepView browser to allow users to check TFBSs. In each species, TFBSs from different transcription factors and many different cell lines were combined and sorted according to their genomic positions. And then, the overlapping TFBSs were grouped into clusters and were imported into database; each cluster included at least one TFBS. Known transcription factor binding matrices were downloaded from the JASPAR (18), Transfac (19), Cistrome (20) and UniPROBE (21) databases.

All of the known lncRNAs or large intergenic non-coding RNAs were downloaded from the Supplementary Data of the six original research articles that addressed these RNAs (22–27) or extracted from Ensembl (28), refSeq (17) and UCSC Bioinformatics website (17). Known functional lncRNAs were downloaded from lncRNAdb database (29). All of the known miRNAs were downloaded from miRBase [release 17.0, (30)]. miRNA targets were downloaded from starBase database (31). All of the refSeq genes were downloaded from the UCSC bioinformatics websites (17). Other known non-coding RNAs were downloaded from the Ensembl database (28) or the UCSC websites (17) or were obtained from the relevant literature. The human (UCSC hg19), mouse (UCSC mm9, NCBI Build 37), dog (UCSC canFam2), chicken (UCSC galGal3), *D. melanogaster* (UCSC dm3) and *C. elegans* (WS190) genome sequences were downloaded from the UCSC bioinformatics websites (17).

Pre-miRNAs were grouped into transcriptional units. TFs might not almost exclusively bind at proximal

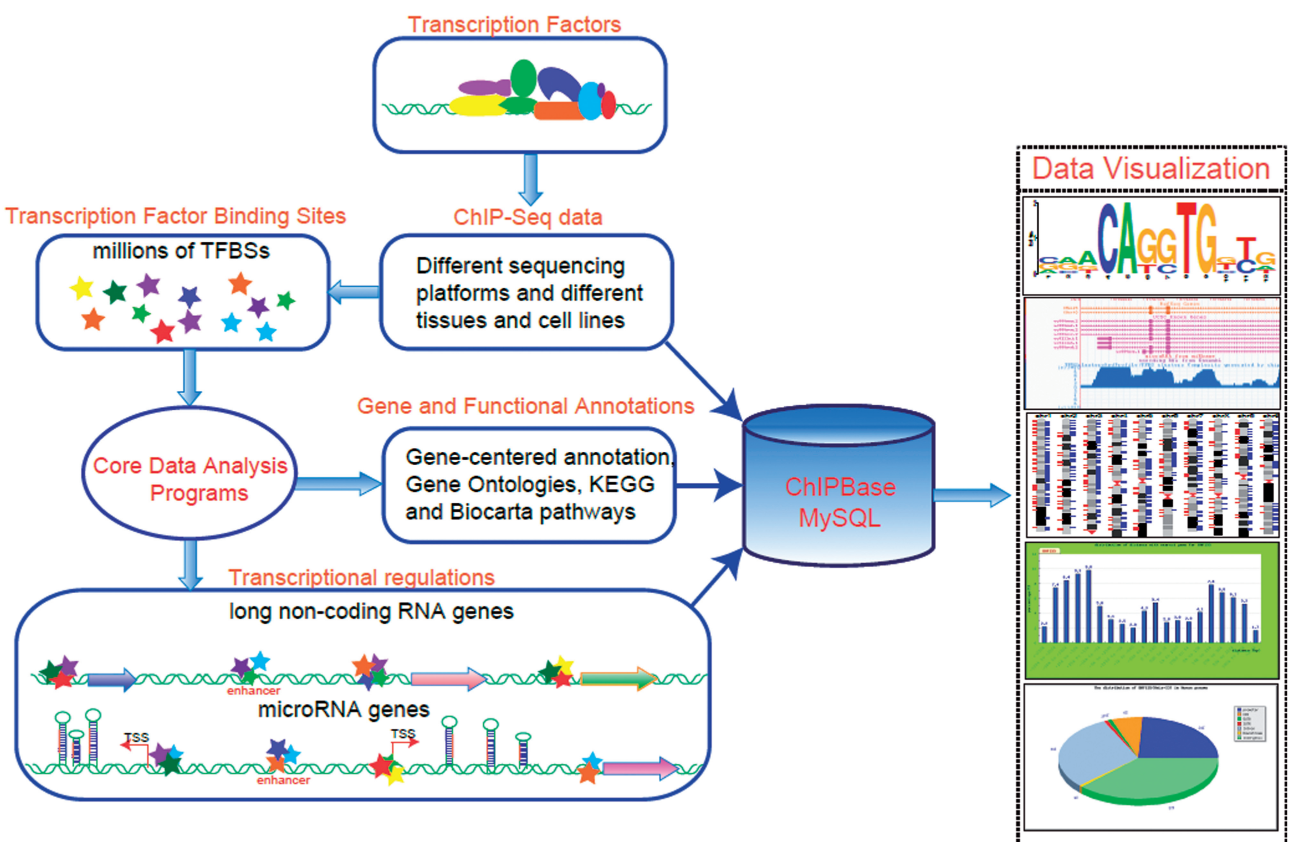

Figure 1. A system-level overview of the core framework of ChIPBase. All of the results that are generated by ChIPBase are deposited in MySQL relational databases and displayed in the visual browser and web page.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

Table 1. The data that are incorporated into ChIPBase

Species	Experiments	TFBSs	TFBS cluster	TF-miRNA	TF-lncRNA
Human	329	6 134 098	1 525 778	41 809	640 741
Mouse	119	2 042 652	478 331	7306	109 383
Dog	2	65 674	52 368	149	/
Chicken	3	40 635	37 108	391	/
D. melanogaster	53	160 202	29 565	1788	98 710
C. elegans	37	216 992	42 666	1790	/

These statistics indicate the numbers of sequencing experiments (ChIP-Seq), TFBSs, TFBS clusters (transcription factor binding clusters), TF-lncRNA interactions and TF-miRNA interactions that are incorporated into ChIPBase. These data are from six organisms, namely, human, mouse, dog, chicken, *D. melanogaster* and *C. elegans*. Known lncRNAs for dog, chicken and *C. elegans* are not available, and therefore the TF-lncRNA interactions for these species are denoted as '/' in the table.

promoters of lncRNAs or miRNAs. For protein-coding genes, more than half of the observed binding events are distal events (32). Moreover, the distances between transcription initiation sites (TSSs) and miRNA genes dramatically vary, ranging from a hundred bases to thousands of bases upstream (33,34). To incorporate proximal and distal binding events (32), for intergenic miRNAs, the 30-kb region upstream and the 10-kb region downstream of the TSS of the first pre-miRNA in the same transcriptional unit were chosen as the regulatory domain (32) of the examined miRNAs. For intronic miRNAs, the 30-kb region upstream and the 10-kb region downstream of TSS of the host genes contained miRNAs were chose as the regulatory domain of the examined miRNAs. The same strategy was used for lncRNAs; we chose a 30-kb region upstream and a 10-kb region downstream of the TSS of each lncRNA as the regulatory domain (32) of each lncRNA. Five-kb upstream region and 1-kb downstream region of each lncRNA and miRNA were chosen as promoter region (32). In each species, regulatory domains/regions of each lncRNA/miRNA were intersected with TFBSs of each data set to identify TFs that regulated the examined ncRNAs. And then, TFBSs overlapping with regulatory domains and corresponding lncRNAs were imported into MySql database.

DATABASE CONTENT

The genome-wide binding maps and high-occupancy target regions of transcription factors

We integrated ~8.7 million TFBSs from 543 ChIP-Seq experiments in various tissues or cell lines to provide comprehensive genome-wide transcription factor binding profiles. To provide more useful information, we generated extensive annotations and analyses for transcription factors and TFBSs. For each TFBS, we identified the nearest/target gene and the distance between the site and the gene, as well as the expression pattern of the TF and its target gene in various tissues or cell lines (Supplementary Figure S1). For each ChIP-Seq experiment, we identified the distribution of TFBSs in the body of the gene and the distribution of the distances of the TFBSs that are associated with the TSSs of the nearest genes, and we provided descriptions of the ChIP-Seq experiments and the expression patterns of the TFs (Supplementary Figure S2). In addition, we offered the chipGO and chipKEGG tools to explore the features of the lists of TF-target interactions that are derived from the ChIP-Seq data (Figure 1 and Supplementary Figure S3).

To directly investigate potential high-occupancy target (HOT) regions on a genome-wide scale, we grouped ~8.7 million TFBSs into ~2 million clusters (Table 1). Each cluster contains between 1 and 74 transcription factors. We designated the genomic locations that were bound by many TFs as HOT regions. For instance, we identified 26 664 HOT regions that were bound by $\geq$15 factors in the human genome. In addition, we generated distribution maps of the numbers of transcription factors in the clusters. The maps are presented in the form of cluster peaks, which are displayed in our deepView genome browser (Figure 1 and Supplementary Figure S4). This display method allows us easily to determine HOT regions of TFs. We also identified the nearest/target genes of these clusters and created a web interface to display this information (Figure 1 and Supplementary Figure S4).

The annotation and identification of TF-lncRNA and TF-miRNA regulatory relationships

To investigate TF–lncRNA and TF–miRNA regulatory relationships, the regulatory domains (see the Materials and Methods section for a detailed description) of lncRNAs and miRNAs were intersected with all TFBSs from diverse tissues and cell lines. In total, we identified ~848 834 TF-lncRNA regulatory relationships between 221 TFs and 38 293 lncRNA transcripts, as well as 53 233 TF-miRNA regulatory relationships between 249 TFs and 2294 miRNA clusters (Table 1). Because of its integration of the large number of high-resolution ChIP-Seq data from diverse tissues and cell lines, this analysis provides an enhanced resolution of these regulatory relationships. Moreover, to enable us to explore the interplay between miRNA transcriptional and posttranscriptional regulation, we integrated the targets of miRNAs from our starBase database (31) into the TF-miRNA networks. Cytoscape (web version) (35) were used to display and draw the TF-miRNA and miRNA-target networks.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

WEB INTERFACE

The use of the deepView and genomeView genome browsers for the comparative analysis of TFBSs and the TF–lncRNA or TF–miRNA regulatory relationships

The large quantity of TFBSs and high-throughput ChIP-Seq data has increased the demand for visual tools that allow for the rapid visual correlation of different types of information. To enable the user to browse seamlessly along the genome and to zoom effortlessly in a very large set of ChIP-Seq data, our improved deepView genome browser (31,36) was developed. This browser provides an integrated view of TFBSs, lncRNAs, miRNAs, protein-coding genes, TF cluster peaks and TF clusters (Figure 2). In the deepView genome browser, the 'zoom out' or 'zoom in' button can be used to extend or shrink the width of the displayed coordinate range. A click on a track item (e.g. a miRNA, lncRNA or TFBS) of interest launches a multiple-alignment trace viewer that displays all of the traces that are relevant to the item in question or links to external resources, such as NCBI, UCSC and miRBase, that can be used to obtain more comprehensive information.

To provide the whole-genome-scale visualization of large-scale TFBSs, miRNAs and lncRNAs, a new genome browser, genomeView, was developed in this study (Figure 3). The user of this browser can view data for a single ChIP-Seq experiment across the entire genome in the context of miRNAs or lncRNAs. TFBSs and miRNAs or lncRNAs are displayed for each location in the genome as a profile over the chromosome ideogram. This feature allows the user to quickly observe genome-scale patterns in the regulatory data and identify regions of interest for further visualization in our deepView genome browser.

The web-based exploration of TF-lncRNA and TF-miRNA regulatory relationships

We provide two web interfaces, LncRNA and MicroRNA, which may be used to display the TF-lncRNA and TF-miRNA interaction relationships, respectively (Supplementary Figures S5 and S6). Users can browse the relationships by entering a lncRNA name. When a user starts typing a lncRNA name in the search box, suggested lncRNA names are displayed in the list box. The user can then either choose a lncRNA from the list box or finish typing a full gene name. The user can also select a TF and search for lncRNAs that are regulated by the selected TF. If users do not enter

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

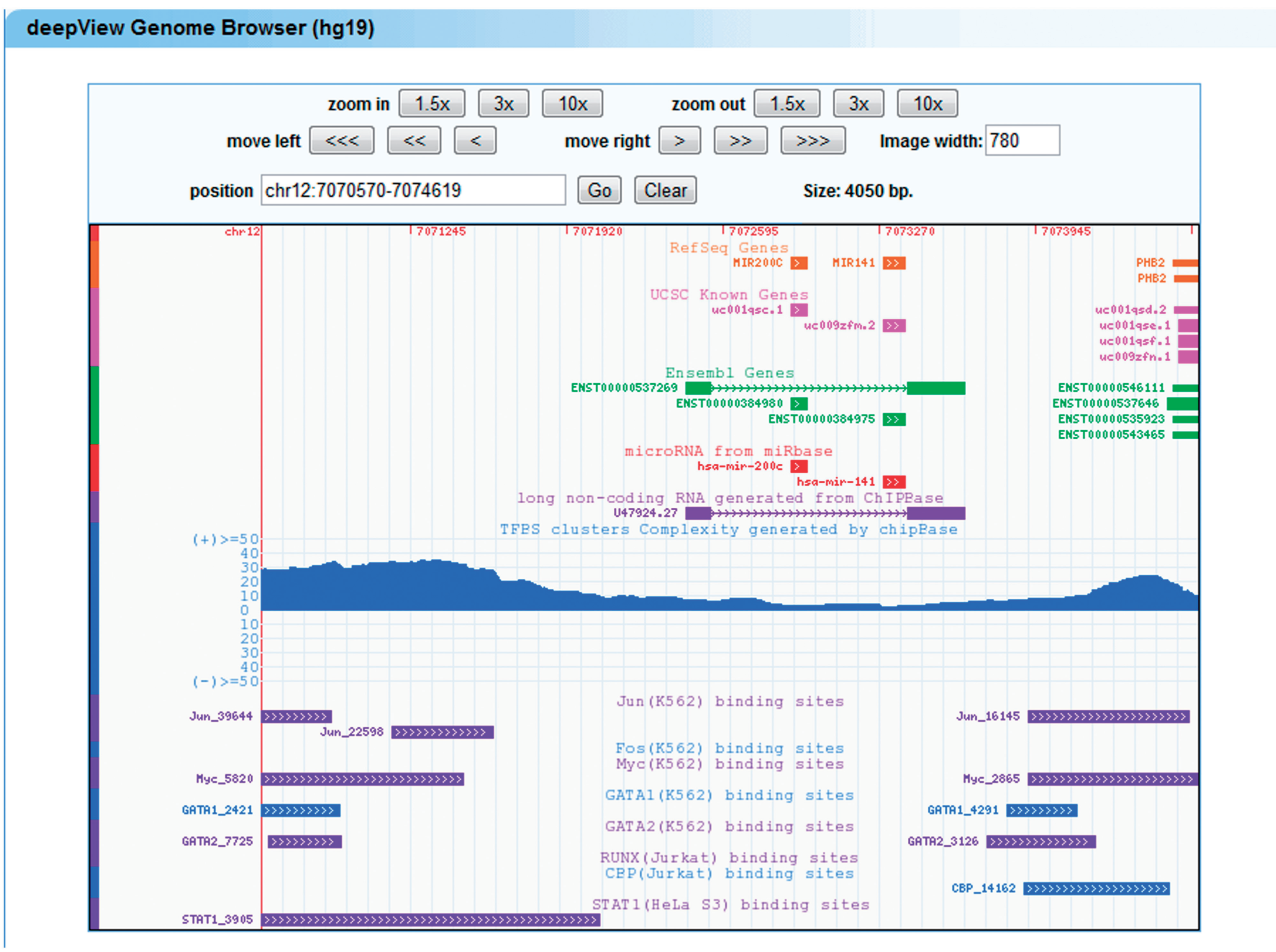

Figure 2. Illustrative screen shots from the deepView browser. The deepView browser provides TFBSs that have been identified from ChIP-Seq data, predicted TFBSs, lncRNAs, miRNAs, protein-coding genes and TFBS clusters.

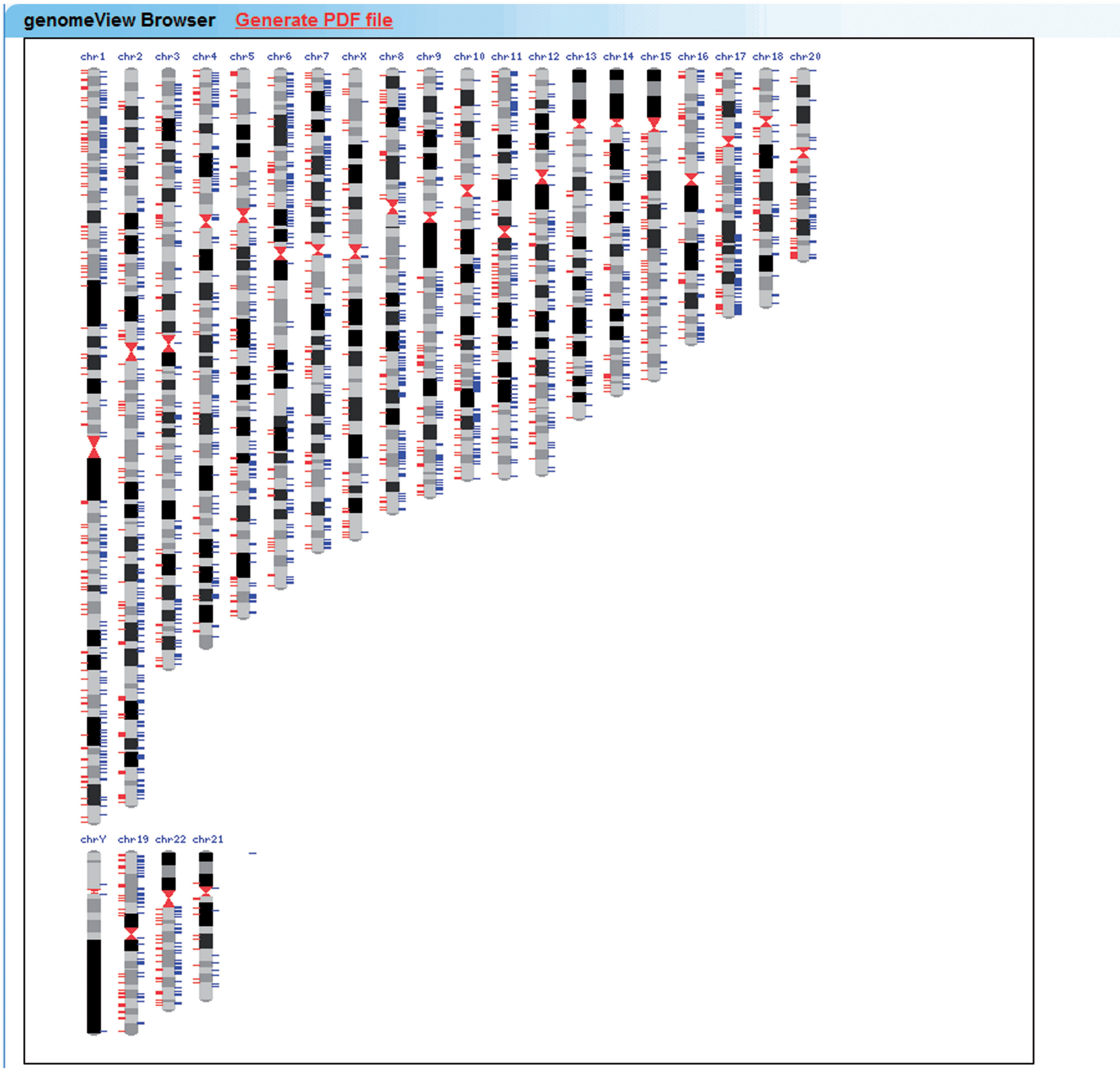

Figure 3. Snapshot of the genomeView browser. The genomeView browser provides the whole-genome-scale visualization of large-scale TFBSs, miRNAs and lncRNAs.

lncRNA name and TF name, webpage will output all the TF-lncRNA interactions. Users can download these interactions to construct more complex networks composed by dozens to hundreds of lncRNAs. The results of the search are listed as the TF-lncRNA table. For the lncRNA interface, the numbers of TFBSs for each lncRNA in are indicated in a table. The users can click on a number within the table to launch a detailed page that provides further information about the TF-lncRNA interaction in question. The user can also click on the title of the table to sort TF-lncRNA interactions according to various features, such as the number of TFBSs, the lncRNA names or the TF names. The detailed information for a TF-lncRNA interaction includes a description of the TF gene and its distance to the start site of the lncRNA (Supplementary Figure S5). The 'references' section enables the retrieval of the primary articles yielding the annotation data. Click the article title link to visit the NCBI PUBMED website.

The microRNA interface is organized similarly to the LncRNA interface. The user can select a miRNA and a TF gene from a drop-down menu to explore TF-miRNA interactions. The numbers of upstream and downstream TFBSs, the genomic coordinates and the distance to the start site of the miRNA are all presented in a table (Supplementary Figure S6A and B). In addition, we have constructed a webpage, Networks, to simply display TF-miRNA, miRNA-target and TF-target interactions (37) using cytoscape (web version) (35) by integrating our starBase database. Users can select different miRNA target regulated by examined miRNA to construct regulatory networks (Supplementary Figure S6C). For example, we recapitulated the published c-Myc, E2F1 and miR-20 network (38) by selecting hsa-miR-20a and miRNA-target gene E2F1 in Networks webpage (Supplementary Figure S6D).

The interface of the transcriptional regulatory for other ncRNAs (such as snoRNAs, tRNAs, snRNAs, etc.) is also

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

provided and organized similarly to the lncRNA and miRNA interfaces. Users can explore their regulatory interactions by similar ways.

The web-based annotation of transcription factor binding regions

We also provide the annotatedTool program, which offers a simple and user-friendly interface to annotate transcription factor binding regions (TFBRs). The user is required to select an intended organism and annotated TSSs of known protein-coding genes, lncRNAs or miRNAs and then upload TFBRs in the browser extensible (BED) format. After the user has completed the data submission, a typical iteration of the annotatedTool program may require several minutes to finish. The output of this program consists of three parts: the distribution of distances between the center of the TFBR and the TSS, information about the nearest gene and a link to the deepView genome browser, which allows the user to view various features of each target region (Supplementary Figure S7).

EXAMPLE APPLICATIONS

In the following section, we will present several example applications of ChIPBase.

hsa-miR-122, a target of liver-enriched transcription factors

Let us assume that we are interested in liver-specific TFs as transcriptional regulators of miR-122, which is also expressed in the liver. We select three liver-enriched TFs (HNF4A, CEBPA and HNF3B/FoxA2) and the miR-122 gene in the microRNA webpage. The results page summarizes all of the query results: (i) there are five ChIP-Seq experiments for these three TFs (Supplementary Figure S8A and B); (ii) there are HNF4A ChIP-Seq data from three different experiments; and (iii) HNF4A and HNF3B/FoxA2 have multiple binding sites in the regulatory domain of miR-122 (Supplementary Figure S8A and B). We navigate to the corresponding deepView genome browser by locating the miR-122 regulatory domain, which opens up a genome browser view that effectively recapitulates the published TFBSs (39) (Figure 4).

lncRNAs as targets of embryonic stem cell transcription factors

To relate a lncRNA gene to the core transcriptional circuitry of embryonic stem (ES) cells, we select nine pluripotency-associated transcription factors (including Oct4, Sox2, Nanog, c-Myc, n-Myc, Klf4, Zfx, E2F1 and Smad1) in the mouse genome at the lncRNA webpage. The results page summarizes the pluripotency-associated transcription factors that bind in the regulatory domains of lncRNAs. A click on linc1428, a known ES-cells-associated lncRNA, launches a deepView genome browser view that also recapitulates the published TFBSs of E2F1, n-Myc and Klf4 (25) (Figure 5).

DISCUSSIONS AND CONCLUSIONS

Ultra-high-throughput next-generation sequencing technology has recently been developed for mapping TFBSs (10,11). In this study, we performed a large-scale integration of public TFBSs that have been generated by high-throughput ChIP-Seq technology and provide the most comprehensive TF data set for various cell types that are available at the present time. We also provide comprehensive transcriptional regulatory maps of lncRNAs and miRNAs by connecting TFs to these non-coding genes.

The transcriptional regulation of the majority of miRNAs and almost all of the discovered lncRNAs is currently unknown. Recent studies have revealed that the deregulation of miRNAs and lncRNAs is correlated with various human cancers and diseases (6,7), and that this deregulation is often due to the aberrant expression of TFs (39). In the current study, we developed the ChIPBase database to decode the transcriptional regulation of lncRNA and miRNA genes from ChIP-Seq data. We can use ChIPBase to recapitulate the known transcriptional regulatory relationships of miRNAs and lncRNAs. For example, ChIPBase can be used to identify that the liver-specific miR-122 is regulated by three liver-enriched TFs (39), and that the linc2048 lncRNA is regulated by embryonic stem cell (ESC)-associated transcription factors (25). In addition, the integration of a large quantity of ChIP-Seq data from diverse tissues and cell lines allows us to provide enhanced resolution and novel findings.

In comparison with other sources, for elucidating the transcriptional regulation of lncRNAs and miRNAs, or storing and analyzing ChIP-Seq data, the distinctive features in our ChIPBase database are as follows. (i) Our ChIPBase database is the first database that provides the transcriptional regulation maps for lncRNA genes. (ii) The other databases that are related to transcriptional regulation for miRNAs, including transmiR (40) and CircuitsDB (41), only collect computationally predicted or experimentally supported TF-miRNA interactions. By contrast, ChIPBase provides the comprehensive TF-miRNA regulatory relationships that have been identified from high-throughput ChIP-Seq data. The entries in TransmiR database contain only the name of TFs and their corresponding target miRNAs. The detail information of TFBSs and TFs, however, is not included. Also, the TransmiR database may not contain the comprehensive target miRNAs of the corresponding TFs. We used two TFs, E2F1 and MYC, whose target miRNAs are the most comprehensive in TransmiR to perform comparison between TransmiR database and our ChIPBase database. When considered only the relationships of TFs and target miRNAs, ChIPBase could identify 77% (20/26) E2F1–miRNA and 75% (21/28) MYC–miRNA relationships. These data indicated that ChIPBase could recover majority of TF-miRNA interactions documented in TransmiR. Moreover, our database also identifies tens of novel E2F1–miRNA and MYC–miRNA relationships that were not included in transmiR data. In addition, TransmiR does not contain HN4FA-miR-122 and CEBPA-miR-122 interactions described in our Example

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

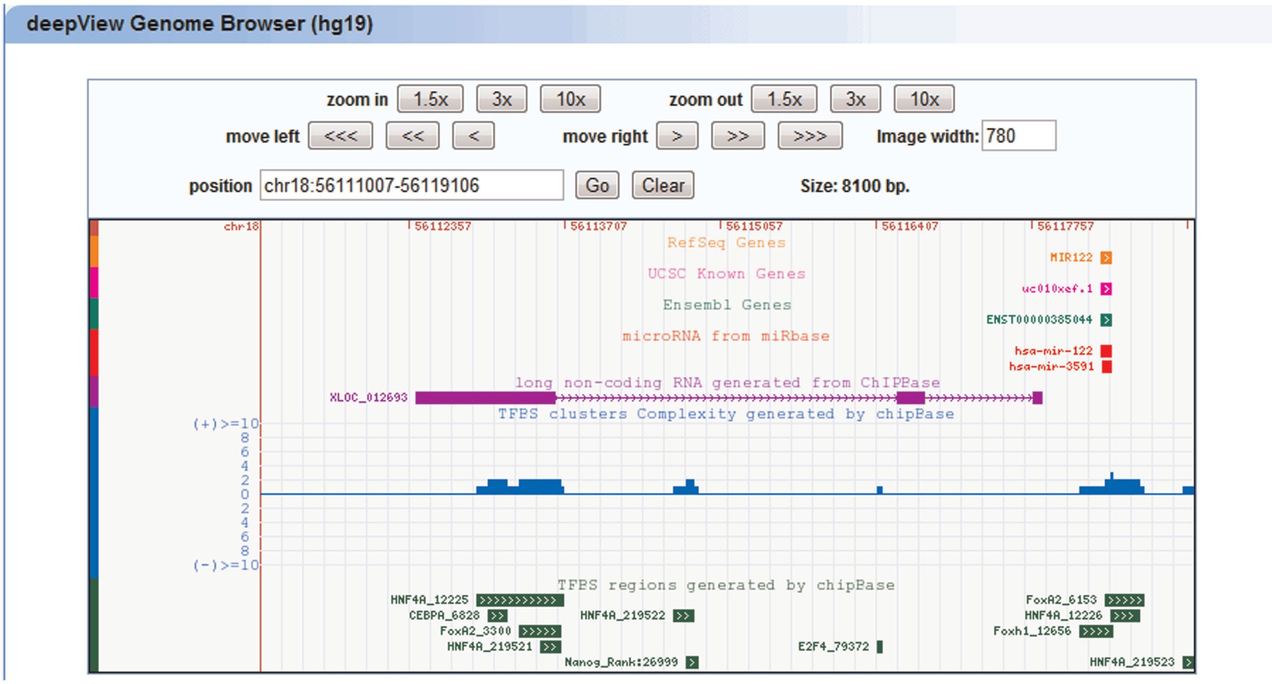

Figure 4. Liver-enriched transcription factors regulate the expression of hsa-miR-122. The liver-enriched transcription factors (HNF4A, CEBPA and HNF3B/FoxA2) bind to the promoter regions of the primary transcript (XLOC_012693) of hsa-miR-122.

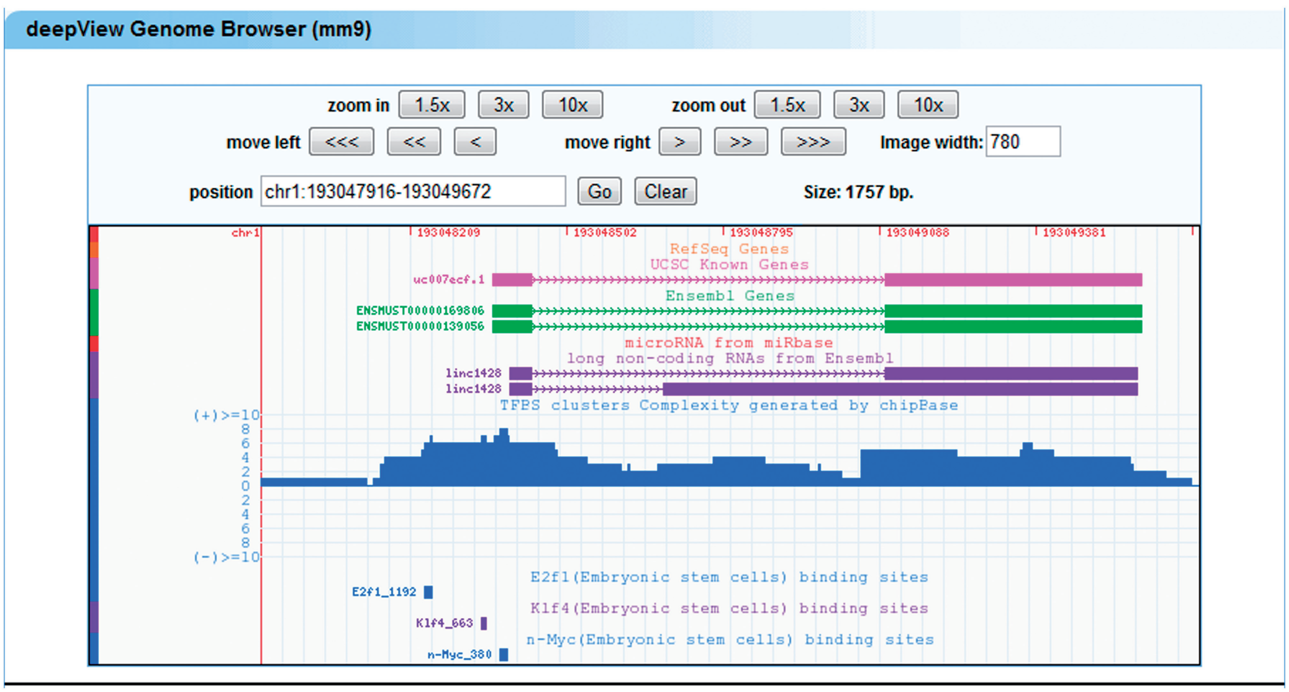

Figure 5. linc1428 as a target of ES cell transcription factors. The ES cell transcription factors (E2F1, n-Myc and Klf4) bind to promoter regions of the linc1428 lncRNA.

Applications sections. (iii) hmChIP (42) is a database of genome-wide ChIP data in human (hg18 version) and mouse (mm8 version). It just provides the Protein–DNA binding intensities from individual samples for user-provided genomic regions. Currently, hmChIP does not explore the transcriptional regulatory of lncRNAs or miRNAs, even for protein-coding genes. By contrast, our ChIPBase provides comprehensive transcriptional regulatory relationships of lncRNAs, miRNAs and other ncRNAs, as well as comprehensive annotation of

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

TFBSs from 543 ChIP-Seq data sets from six organisms. (iv) To enable the user to browse seamlessly along the genome and to zoom effortlessly in a very large set of ChIP-Seq data, our improved deepView genome browser was developed to provide an integrated view of TFBSs that have been identified from ChIP-Seq data, predicted TFBSs, ncRNAs, protein-coding genes and TFBS clusters (Figures 2 and 3). (v) We developed two web tools, annotatedTool and genomeViewer to annotate and discover the transcriptional regulatory relationships of lncRNAs and miRNAs from ChIP-Seq data (Figure 1). (vi) We constructed genome-wide transcription factor binding profiles from ChIP-Seq data. Combinatorial transcription factor interactions that control the transcriptional regulations of lncRNAs and miRNAs were easily identified by searching for appropriate profiles in the genome browser (Figure 2). (vii) ChIPBase also provides the gene ontology annotation, biological pathways and expression patterns of transcription factor binding targets (Figure 1). This supplementary information may provide valuable insights into the function of each TF, lncRNA and miRNA. Finally, the data and the integrative, interactive and versatile displays that are provided by the ChIPBase database will aid future experimental and computational studies in their attempts to elucidate the regulation of lncRNAs and miRNAs by TFs and assess the roles of these regulatory relationships in human diseases.

FUTURE DIRECTIONS

As a means of comprehensively integrating ChIP-Seq data, ChIPBase is expected to provide considerable resources to assist researchers that are investigating the TF-lncRNA and TF-miRNA regulatory networks and examining the biological functions of the genes and ncRNAs with expression levels that are controlled by transcription factors. As ChIP-Seq technology is applied to a broader set of species, cell lines, tissues and conditions, ChIPBase will continue to be developed and refined toward the achievement of the following goals: (i) the better integration and cross-comparison of diverse ChIP-Seq data sets and data resources. (ii) the correlation of these diverse ChIP-Seq data with lncRNAs and miRNAs. (iii) we will continue to extend the amount of storage space and improve the performance of our computer servers for storing and analysing these new data, and improve the database to accept upload of new data by the users. In addition, we intend to integrate the epigenomic data that are generated by ChIP-Seq technology into ChIPBase to improve our understanding of eukaryotic regulatory networks.

AVAILABILITY

ChIPBase is freely available at http://deepbase.sysu.edu.cn/chipbase/. The ChIPBase data files can be freely downloaded and used in accordance with the GNU Public License.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1, Supplementary Figures 1–8 and Supplementary References [15,16,43–98].

FUNDING

Ministry of Science and Technology of China, National Basic Research Program [No. 2011CB811300]; National Natural Science Foundation of China [No. 31230042, 30900820, 81070589]; funds from Guangdong Province [No. S2012010010510]; The project of Science and Technology New Star in ZhuJiang Guangzhou city [No. 2012J2200025]; Fundamental Research Funds for the Central Universities [No. 2011330003161070]; China Postdoctoral Science Foundation [No. 200902348]; Guangdong Province Key Laboratory of Computational Science and the Guangdong Province Computational Science Innovative Research Team (in part). Funding for open access charge: Ministry of Science and Technology of China, National Basic Research Program [No. 2011CB811300].

Conflict of interest statement. None declared.

REFERENCES

1. Filipowicz,W., Bhattacharyya,S.N. and Sonenberg,N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, **9**, 102–114.
2. Inui,M., Martello,G. and Piccolo,S. (2010) MicroRNA control of signal transduction. *Nat. Rev. Mol. Cell Biol.*, **11**, 252–263.
3. Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Maeda,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
4. Mercer,T.R., Dinger,M.E., Sunkin,S.M., Mehler,M.F. and Mattick,J.S. (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA*, **105**, 716–721.
5. Guttman,M. and Rinn,J.L. (2012) Modular regulatory principles of large non-coding RNAs. *Nature*, **482**, 339–346.
6. Huarte,M. and Rinn,J.L. (2010) Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet.*, **19**, R152–R161.
7. Spizzo,R., Almeida,M.I., Colombatti,A. and Calin,G.A. (2012) Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene*, **31**, 4577–4587.
8. Prensner,J.R. and Chinnaiyan,A.M. (2011) The emergence of lncRNAs in cancer biology. *Cancer Discov.*, **1**, 391–407.
9. Tompa,M., Li,N., Bailey,T.L., Church,G.M., De Moor,B., Eskin,E., Favorov,A.V., Frith,M.C., Fu,Y., Kent,W.J. *et al.* (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.*, **23**, 137–144.
10. Park,P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, **10**, 669–680.
11. Pepke,S., Wold,B. and Mortazavi,A. (2009) Computation for ChIP-seq and RNA-seq studies. *Nat. Methods*, **6**, S22–S32.
12. Farnham,P.J. (2009) Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.*, **10**, 605–616.
13. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M. *et al.* (2011) NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res.*, **39**, D1005–D1010.
14. The ENCODE Project Consortium. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, **9**, e1001046.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

15. Negre,N., Brown,C.D., Ma,L., Bristow,C.A., Miller,S.W., Wagner,U., Kheradpour,P., Eaton,M.L., Loriaux,P., Sealfon,R. *et al*. (2011) A cis-regulatory map of the Drosophila genome. *Nature*, **471**, 527–531.
16. Gerstein,M.B., Lu,Z.J., Van Nostrand,E.L., Cheng,C., Arshinoff,B.I., Liu,T., Yip,K.Y., Robilotto,R., Rechtsteiner,A., Ikegami,K. *et al*. (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science*, **330**, 1775–1787.
17. Dreszer,T.R., Karolchik,D., Zweig,A.S., Hinrichs,A.S., Raney,B.J., Kuhn,R.M., Meyer,L.R., Wong,M., Sloan,C.A., Rosenbloom,K.R. *et al*. (2012) The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res*., **40**, D918–D923.
18. Portales-Casamar,E., Thongjuea,S., Kwon,A.T., Arenillas,D., Zhao,X., Valen,E., Yusuf,D., Lenhard,B., Wasserman,W.W. and Sandelin,A. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res*., **38**, D105–D110.
19. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K. *et al*. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*., **34**, D108–D110.
20. Liu,T., Ortiz,J.A., Taing,L., Meyer,C.A., Lee,B., Zhang,Y., Shin,H., Wong,S.S., Ma,J., Lei,Y. *et al*. (2011) Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol*., **12**, R83.
21. Robasky,K. and Bulyk,M.L. (2011) UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res*., **39**, D124–D128.
22. Cabili,M.N., Trapnell,C., Goff,L., Koziol,M., Tazon-Vega,B., Regev,A. and Rinn,J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*., **25**, 1915–1927.
23. Young,R.S., Marques,A.C., Tibbit,C., Haerty,W., Bassett,A.R., Liu,J.L. and Ponting,C.P. (2012) Identification and properties of 1119 candidate lincRNA loci in the Drosophila melanogaster genome. *Genome Biol*. *Evol*., **4**, 427–442.
24. Guttman,M., Garber,M., Levin,J.Z., Donaghey,J., Robinson,J., Adiconis,X., Fan,L., Koziol,M.J., Gnirke,A., Nusbaum,C. *et al*. (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat*. *Biotechnol*., **28**, 503–510.
25. Guttman,M., Donaghey,J., Carey,B.W., Garber,M., Grenier,J.K., Munson,G., Young,G., Lucas,A.B., Ach,R., Bruhn,L. *et al.* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, **477**, 295–300.
26. Guttman,M., Amit,I., Garber,M., French,C., Lin,M.F., Feldser,D., Huarte,M., Zuk,O., Carey,B.W., Cassady,J.P. *et al*. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
27. Belgard,T.G., Marques,A.C., Oliver,P.L., Abaan,H.O., Sirey,T.M., Hoerder-Suabedissen,A., Garcia-Moreno,F., Molnar,Z., Margulies,E.H. and Ponting,C.P. (2011) A transcriptomic atlas of mouse neocortical layers. *Neuron*, **71**, 605–616.
28. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al*. (2012) Ensembl 2012. *Nucleic Acids Res*., **40**, D84–D90.
29. Amaral,P.P., Clark,M.B., Gascoigne,D.K., Dinger,M.E. and Mattick,J.S. (2011) lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res*., **39**, D146–D151.
30. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*., **39**, D152–D157.
31. Yang,J.H., Li,J.H., Shao,P., Zhou,H., Chen,Y.Q. and Qu,L.H. (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res*, **39**, D202–D209.
32. McLean,C.Y., Bristor,D., Hiller,M., Clarke,S.L., Schaar,B.T., Lowe,C.B., Wenger,A.M. and Bejerano,G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat*. *Biotechnol*., **28**, 495–501.
33. Lee,Y., Kim,M., Han,J., Yeom,K.H., Lee,S., Baek,S.H. and Kim,V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*., **23**, 4051–4060.
34. Chang,T.C., Wentzel,E.A., Kent,O.A., Ramachandran,K., Mullendore,M., Lee,K.H., Feldmann,G., Yamakuchi,M., Ferlito,M., Lowenstein,C.J. *et al*. (2007) Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol*. *Cell*., **26**, 745–752.
35. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*., **13**, 2498–2504.
36. Yang,J.H., Shao,P., Zhou,H., Chen,Y.Q. and Qu,L.H. (2010) deepBase: a database for deeply annotating and mining deep sequencing data. *Nucleic Acids Res*., **38**, D123–D130.
37. Lai,X., Schmitz,U., Gupta,S.K., Bhattacharya,A., Kunz,M., Wolkenhauer,O. and Vera,J. (2012) Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs. *Nucleic Acids Res*., **40**, 8818–8834.
38. O'Donnell,K.A., Wentzel,E.A., Zeller,K.I., Dang,C.V. and Mendell,J.T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, **435**, 839–843.
39. Xu,H., He,J.H., Xiao,Z.D., Zhang,Q.Q., Chen,Y.Q., Zhou,H. and Qu,L.H. (2010) Liver-enriched transcription factors regulate microRNA-122 that targets CUTL1 during liver development. *Hepatology*, **52**, 1431–1442.
40. Wang,J., Lu,M., Qiu,C. and Cui,Q. (2010) TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res*., **38**, D119–D122.
41. Friard,O., Re,A., Taverna,D., De Bortoli,M. and Cora,D. (2010) CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics*, **11**, 435.
42. Chen,L., Wu,G. and Ji,H. (2011) hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data. *Bioinformatics*, **27**, 1447–1448.
43. Ang,Y.S., Tsai,S.Y., Lee,D.F., Monk,J., Su,J., Ratnakumar,K., Ding,J., Ge,Y., Darr,H., Chang,B. *et al*. (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell*, **145**, 183–197.
44. Blow,M.J., McCulley,D.J., Li,Z., Zhang,T., Akiyama,J.A., Holt,A., Plajzer-Frick,I., Shoukry,M., Wright,C., Chen,F. *et al.* (2010) ChIP-Seq identification of weakly conserved heart enhancers. *Nat*. *Genet*., **42**, 806–810.
45. Bottomly,D., Kyler,S.L., McWeeney,S.K. and Yochum,G.S. (2010) Identification of beta-catenin binding regions in colon cancer cells using ChIP-Seq. *Nucleic Acids Res*., **38**, 5735–5745.
46. Bradley,R.K., Li,X.Y., Trapnell,C., Davidson,S., Pachter,L., Chu,H.C., Tonkin,L.A., Biggin,M.D. and Eisen,M.B. (2010) Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related *Drosophila* species. *Plos Biol*, **8**, e1000343.
47. Ender,C., Krek,A., Friedlander,M.R., Beitzinger,M., Weinmann,L., Chen,W., Pfeffer,S., Rajewsky,N. and Meister,G. (2008) A human snoRNA with microRNA-like functions. *Mol*. *Cell*, **32**, 519–528.
48. Cheng,Y., Wu,W., Kumar,S.A., Yu,D., Deng,W., Tripic,T., King,D.C., Chen,K.B., Zhang,Y., Drautz,D. *et al.* (2009) Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res*., **19**, 2172–2184.
49. Cicatiello,L., Mutarelli,M., Grober,O.M.V., Paris,O., Ferraro,L., Ravo,M., Tarallo,R., Luo,S.J., Schroth,G.P., Seifert,M. *et al.* (2010) Estrogen receptor alpha controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs. *Am*. *J*. *Pathol*., **176**, 2113–2130.
50. Frietze,S., Lan,X., Jin,V.X. and Farnham,P.J. (2010) Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *J*. *Biol*. *Chem*., **285**, 1393–1403.
51. Fujiwara,T., O'Geen,H., Keles,S., Blahnik,K., Linnemann,A.K., Kang,Y.A., Choi,K., Farnham,P.J. and Bresnick,E.H. (2009) Discovering hematopoietic mechanisms through genome-wide

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

analysis of GATA factor chromatin occupancy. *Mol. Cell*, **36**, 667–681.
52. Grober,O.M., Mutarelli,M., Giurato,G., Ravo,M., Cicatiello,L., De Filippo,M.R., Ferraro,L., Nassa,G., Papa,M.F., Paris,O. *et al.* (2011) Global analysis of estrogen receptor beta binding to breast cancer cell genome reveals an extensive interplay with estrogen receptor alpha for target gene regulation. *BMC Genomics*, **12**, 36.
53. He,A., Kong,S.W., Ma,Q. and Pu,W.T. (2011) Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc. Natl Acad. Sci. USA*, **108**, 5632–5637.
54. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
55. Heng,J.C.D., Feng,B., Han,J.Y., Jiang,J.M., Kraus,P., Ng,J.H., Orlov,Y.L., Huss,M., Yang,L., Lufkin,T. *et al.* (2010) The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell*, **6**, 167–174.
56. Hollenhorst,P.C., Chandler,K.J., Poulsen,R.L., Johnson,W.E., Speck,N.A. and Graves,B.J. (2009) DNA specificity determinants associate with distinct transcription factor functions. *Plos Genet.*, **5**, e1000778.
57. Hu,M., Yu,J.D., Taylor,J.M.G., Chinnaiyan,A.M. and Qin,Z.H.S. (2010) On the detection and refinement of transcription factor binding sites using ChIP-Seq data. *Nucleic Acids Res.*, **38**, 2154–2167.
58. Kang,J.S., Gemberling,M., Nakamura,M., Whitby,F.G., Handa,H., Fairbrother,W.G. and Tantin,D. (2009) A general mechanism for transcription regulation by Oct1 and Oct4 in response to genotoxic and oxidative stress. *Genes Dev.*, **23**, 208–222.
59. Kim,S.W., Yoon,S.J., Chuong,E., Oyolu,C., Wills,A.E., Gupta,R. and Baker,J. (2011) Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.*, **357**, 492–504.
60. Kleine-Kohlbrecher,D., Christensen,J., Vandamme,J., Abarrategui,I., Bak,M., Tommerup,N., Shi,X., Gozani,O., Rappsilber,J., Salcini,A.E. *et al.* (2010) A functional link between the histone demethylase PHF8 and the transcription factor ZNF711 in X-linked mental retardation. *Mol. Cell*, **38**, 165–178.
61. Kouwenhoven,E.N., van Heeringen,S.J., Tena,J.J., Oti,M., Dutilh,B.E., Alonso,M.E., de la Calle-Mustienes,E., Smeenk,L., Rinne,T., Parsaulian,L. *et al.* (2010) Genome-wide profiling of p63 DNA-binding sites identifies an element that regulates gene expression during limb development in the 7q21 SHFM1 locus. *Plos Genet.*, **6**, e1001065.
62. Kunarso,G., Chia,N.Y., Jeyakani,J., Hwang,C., Lu,X.Y., Chan,Y.S., Ng,H.H. and Bourque,G. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.*, **42**, 631–634.
63. Lee,B.K., Bhinge,A.A. and Iyer,V.R. (2011) Wide-ranging functions of E2F4 in transcriptional activation and repression revealed by genome-wide analysis. *Nucleic Acids Res.*, **39**, 3558–3573.
64. Lefterova,M.I., Steger,D.J., Zhuo,D., Qatanani,M., Mullican,S.E., Tuteja,G., Manduchi,E., Grant,G.R. and Lazar,M.A. (2010) Cell-specific determinants of peroxisome proliferator-activated receptor gamma function in adipocytes and macrophages. *Mol. Cell. Biol.*, **30**, 2078–2089.
65. Lin,Y.C., Jhunjhunwala,S., Benner,C., Heinz,S., Welinder,E., Mansson,R., Sigvardsson,M., Hagman,J., Espinoza,C.A., Dutkowski,J. *et al.* (2010) A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.*, **11**, 635–643.
66. Lo,K.A., Bauchmann,M.K., Baumann,A.P., Donahue,C.J., Thiede,M.A., Hayes,L.S., des Etages,S.A. and Fraenkel,E. (2011) Genome-wide profiling of H3K56 acetylation and transcription factor binding sites in human adipocytes. *PLoS One*, **6**, e19778.
67. MacIsaac,K.D., Lo,K.A., Gordon,W., Motola,S., Mazor,T. and Fraenkel,E. (2010) A quantitative model of transcriptional regulation reveals the influence of binding location on expression. *PLoS Comput. Biol.*, **6**, e1000773.
68. Martin,D., Pantoja,C., Minan,A.F., Valdes-Quezada,C., Molto,E., Matesanz,F., Bogdanovic,O., de la Calle-Mustienes,E., Dominguez,O., Taher,L. *et al.* (2011) Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes (vol 18, pg 708, 2011). *Nat. Struct. Mol. Biol.*, **18**, 1084–1084.
69. Miranda-Carboni,G.A., Guemes,M., Bailey,S., Anaya,E., Corselli,M., Peault,B. and Krum,S.A. (2011) GATA4 regulates estrogen receptor-alpha-mediated osteoblast transcription. *Mol. Endocrinol.*, **25**, 1126–1136.
70. Nitzsche,A., Paszkowski-Rogacz,M., Matarese,F., Janssen-Megens,E.M., Hubner,N.C., Schulz,H., de Vries,I., Ding,L., Huebner,N., Mann,M. *et al.* (2011) RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS One*, **6**, e19470.
71. Pasini,D., Cloos,P.A., Walfridsson,J., Olsson,L., Bukowski,J.P., Johansen,J.V., Bak,M., Tommerup,N., Rappsilber,J. and Helin,K. (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature*, **464**, 306–310.
72. Peng,J.C., Valouev,A., Swigut,T., Zhang,J., Zhao,Y., Sidow,A. and Wysocka,J. (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell*, **139**, 1290–1302.
73. Raha,D., Wang,Z., Moqtaderi,Z., Wu,L., Zhong,G., Gerstein,M., Struhl,K. and Snyder,M. (2010) Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc. Natl Acad. Sci. USA*, **107**, 3639–3644.
74. Robertson,G., Hirst,M., Bainbridge,M., Bilenky,M., Zhao,Y.J., Zeng,T., Euskirchen,G., Bernier,B., Varhol,R., Delaney,A. *et al.* (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods*, **4**, 651–657.
75. Ross-Innes,C.S., Stark,R., Holmes,K.A., Schmidt,D., Spyrou,C., Russell,R., Massie,C.E., Vowler,S.L., Eldridge,M. and Carroll,J.S. (2010) Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. *Genes Dev.*, **24**, 171–182.
76. Rozowsky,J., Euskirchen,G., Auerbach,R.K., Zhang,Z.D.D., Gibson,T., Bjornson,R., Carriero,N., Snyder,M. and Gerstein,M.B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat. Biotechnol.*, **27**, 66–75.
77. Schlesinger,J., Schueler,M., Grunert,M., Fischer,J.J., Zhang,Q., Krueger,T., Lange,M., Tonjes,M., Dunkel,I. and Sperling,S.R. (2011) The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. *Plos Genet.*, **7**, e1001313.
78. Schmidt,D., Wilson,M.D., Ballester,B., Schwalie,P.C., Brown,G.D., Marshall,A., Kutter,C., Watt,S., Martinez-Jimenez,C.P., Mackay,S. *et al.* (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, **328**, 1036–1040.
79. Soccio,R.E., Tuteja,G., Everett,L.J., Li,Z., Lazar,M.A. and Kaestner,K.H. (2011) Species-specific strategies underlying conserved functions of metabolic transcription factors. *Mol. Endocrinol.*, **25**, 694–706.
80. Steger,D.J., Grant,G.R., Schupp,M., Tomaru,T., Lefterova,M.I., Schug,J., Manduchi,E., Stoeckert,C.J. and Lazar,M.A. (2010) Propagation of adipogenic signals through an epigenomic transition state. *Genes Dev.*, **24**, 1035–1044.
81. Stender,J.D., Kim,K., Charn,T.H., Komm,B., Chang,K.C., Kraus,W.L., Benner,C., Glass,C.K. and Katzenellenbogen,B.S. (2010) Genome-wide analysis of estrogen receptor alpha DNA binding and tethering mechanisms identifies Runx1 as a novel tethering factor in receptor-mediated transcriptional activation. *Mol. Cell. Biol.*, **30**, 3943–3955.
82. Treiber,T., Mandel,E.M., Pott,S., Gyory,I., Firner,S., Liu,E.T. and Grosschedl,R. (2010) Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription-independent poising of chromatin. *Immunity*, **32**, 714–725.
83. Verzi,M.P., Shin,H., He,H.H., Sulahian,R., Meyer,C.A., Montgomery,R.K., Fleet,J.C., Brown,M., Liu,X.S. and Shivdasani,R.A. (2010) Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015

for the intestinal transcription factor CDX2. *Dev. Cell*, **19**, 713–726.
84. Verzi,M.P., Shin,H., Ho,L.L., Liu,X.S. and Shivdasani,R.A. (2011) Essential and redundant functions of caudal family proteins in activating adult intestinal genes. *Mol. Cell. Biol.*, **31**, 2026–2039.
85. Visel,A., Blow,M.J., Li,Z., Zhang,T., Akiyama,J.A., Holt,A., Plajzer-Frick,I., Shoukry,M., Wright,C., Chen,F. *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**, 854–858.
86. Vivar,O.I., Zhao,X.Y., Saunier,E.F., Griffin,C., Mayba,O.S., Tagliaferri,M., Cohen,I., Speed,T.P. and Leitman,D.C. (2010) Estrogen receptor beta binds to and regulates three distinct classes of target genes. *J. Biol. Chem.*, **285**, 22059–22066.
87. Walker,E., Chang,W.Y., Hunkapiller,J., Cagney,G., Garcha,K., Torchia,J., Krogan,N.J., Reiter,J.F. and Stanford,W.L. (2010) Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell Stem Cell*, **6**, 153–166.
88. Wang,D., Garcia-Bassets,I., Benner,C., Li,W., Su,X., Zhou,Y., Qiu,J., Liu,W., Kaikkonen,M.U., Ohgi,K.A. *et al.* (2011) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, **474**, 390–394.
89. Warnatz,H.J., Schmidt,D., Manke,T., Piccini,I., Sultan,M., Borodina,T., Balzereit,D., Wruck,W., Soldatov,A., Vingron,M. *et al.* (2011) The BTB and CNC homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle. *J. Biol. Chem.*, **286**, 23521–23532.
90. Wei,G.H., Badis,G., Berger,M.F., Kivioja,T., Palin,K., Enge,M., Bonke,M., Jolma,A., Varjosalo,M., Gehrke,A.R. *et al.* (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and *in vivo*. *EMBO J.*, **29**, 2147–2160.
91. Welboren,W.J., van Driel,M.A., Janssen-Megens,E.M., van Heeringen,S.J., Sweep,F.C., Span,P.N. and Stunnenberg,H.G. (2009) ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.*, **28**, 1418–1428.
92. Wilson,N.K., Foster,S.D., Wang,X., Knezevic,K., Schutte,J., Kaimakis,P., Chilarska,P.M., Kinston,S., Ouwehand,W.H., Dzierzak,E. *et al.* (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, **7**, 532–544.
93. Yao,H., Brick,K., Evrard,Y., Xiao,T., Camerini-Otero,R.D. and Felsenfeld,G. (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.*, **24**, 2543–2555.
94. Yu,H.B., Johnson,R., Kunarso,G. and Stanton,L.W. (2011) Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells. *Genome Res.*, **21**, 1284–1293.
95. Yu,J.D., Yu,J.J., Mani,R.S., Cao,Q., Brenner,C.J., Cao,X.H., Wang,X.J., Wu,L.T., Li,J., Hu,M. *et al.* (2010) An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell*, **17**, 443–454.
96. Yu,M., Riva,L., Xie,H.F., Schindler,Y., Moran,T.B., Cheng,Y., Yu,D.N., Hardison,R., Weiss,M.J., Orkin,S.H. *et al.* (2009) Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol. Cell*, **36**, 682–695.
97. Zhao,L., Glazov,E.A., Pattabiraman,D.R., Al-Owaidi,F., Zhang,P., Brown,M.A., Leo,P.J. and Gonda,T.J. (2011) Integrated genome-wide chromatin occupancy and expression analyses identify key myeloid pro-differentiation transcription factors repressed by Myb. *Nucleic Acids Res.*, **39**, 4664–4679.
98. Zhong,M., Niu,W., Lu,Z.J., Sarov,M., Murray,J.I., Janette,J., Raha,D., Sheaffer,K.L., Lam,H.Y., Preston,E. *et al.* (2010) Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. *Plos Genet.*, **6**, e1000848.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 7, 2015